on the economics of data and privacy

rohit lamba
penn state

feb 21 2020

privacy workshop
princeton & luohan academy
luohan academy report

➤ excellent overview of fundamentals underlying data & privacy.

➤ basic structure

  ➤ understanding the demand for privacy
  ➤ what is the value of data?
  ➤ how is data different than bananas or accountants?
  ➤ how should we regulate the data economy?

➤ insightful experiments conducted over a large sample.
implications for regulation:
- should Facebook have been allowed to buy Instagram?
- should Google/Ali Baba be (global) data hubs?

implications for modeling:
- should ‘we’ think about asymmetric information differently?
- an example using insurance markets.

hey, but what is privacy?
- is it a type or is it a social norm?
- is it intrinsic or instrumental?
a tale of two takeovers

- AT&T wanted to buy T-Mobile in 2011 for $39 billion.
  - DOJ challenged it, AT&T abandoned the takeover.
  - Cons: higher prices, fewer choices, and less innovation,
  - Pros: increase network quality and improve balance sheets.

- FCC doc on the proposed merger:

  Commission staff finds that the applicants have failed... proving that the proposed transaction, on balance, will serve the public interest. Upon careful examination... the staff concludes that significant harms to competition are likely to result, primarily in the form of increased prices for consumers, reduced incentives for innovation, an decreased consumer choice...
a tale of two takeovers

- Facebook bought Instagram in 2012 for $1b
  - Facebook was desktop oriented, Instagram was a mobile app
  - No major regulatory blockade.
  - In 2019, Instagram more than 1/4th of Facebook's revenue
  - Crudely makes Instagram's market cap $150b this morning!

- What is the tradeoff?
  - Maybe consumer is better off matching data across platforms,
  - But that makes Facebook a large monopoly.

- Argue that the "transaction" doesn't fit the neo-class model.
**facebook & google: the kill zone**

<table>
<thead>
<tr>
<th>year</th>
<th>parent</th>
<th>target</th>
<th>price in $b</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>google</td>
<td>youtube</td>
<td>1.65</td>
</tr>
<tr>
<td>2007</td>
<td>google</td>
<td>doubleclick</td>
<td>3.1</td>
</tr>
<tr>
<td>2009</td>
<td>google</td>
<td>admob</td>
<td>0.75</td>
</tr>
<tr>
<td>2009</td>
<td>google</td>
<td>postini</td>
<td>0.625</td>
</tr>
<tr>
<td>2011</td>
<td>google</td>
<td>ita software</td>
<td>0.67</td>
</tr>
<tr>
<td>2012</td>
<td>facebook</td>
<td>instragram</td>
<td>1</td>
</tr>
<tr>
<td>2013</td>
<td>google</td>
<td>waze</td>
<td>0.96</td>
</tr>
<tr>
<td>2014</td>
<td>facebook</td>
<td>whatsapp</td>
<td>19</td>
</tr>
<tr>
<td>2016</td>
<td>google</td>
<td>apigee</td>
<td>0.625</td>
</tr>
</tbody>
</table>

source: pitchbook, and kamepalli-rajan-zingales [2019]
let’s theorize a bit

- how is facebook’s takeover of Instagram different?
  - two-sided platform, one faces advertisers, and other consumers,
  - the consumer side is priced at zero,
  - huge network effects on the consumer side, and
  - switching costs.

- ‘standard’ economics forces kaput?
  - how to measure “competition” in the absence of prices?
  - network effects push towards consolidation anyway?

- since “prices” continue to be zero, what’s problem anyway?
  - maybe not internalizing the upstream/downstream markets?
  - innovation still a big issue, schumpeter turning in the grave.

- the politics of the day: brake ’em up!
what is special about the data economy?

▶ are data monopolies harder to regulate?
  ▶ access to big data makes firms get constantly better
  ▶ network effects are strong: more valuable with more customers.

▶ at least three issues:
  ▶ evaluating efficiency, concentration, etc.
  ▶ quantifying price discrimination,
  ▶ privacy: what part of the data is being used and for what?

▶ perhaps standard approaches of break 'em up less desirable?

▶ is the solution (inter)national data repositories?
  ▶ designing newer anti-trust laws,
  ▶ partitioning data interesting computational and social problem.
what does this mean for modeling?

- all of this means an exciting timing for economists.
- since we are at drawing board stage (and 'data' is scare)
  - a good time to explore various models
  - a great opportunity for theory!
insurance markets where firm has greater statistical info.

basic setup same as rothschild-stiglitz [1976].

“state of the world” is 2-d: $\theta = (\theta_1, \theta_2)$ s.t. $\theta_i \in \{H, L\}$.

extent of damage is given by $\mu_{ij}$ s.t.

- $\mu_{HH} > \mu_{HL} > \mu_{LL}$,
- $\mu_{HH} > \mu_{LH} > \mu_{LL}$.

joint distribution of $\theta$ is given by $q = \{q_{HH}, q_{HL}, q_{LH}, q_{LL}\}$. 
<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$L$</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$q_{LL}$</td>
<td>$q_{LH}$</td>
</tr>
<tr>
<td>$H$</td>
<td>$q_{HL}$</td>
<td>$q_{HH}$</td>
</tr>
</tbody>
</table>

$\theta_2$

$q_1$

$q_2$

$1 - q_2$

$1 - q_1$


<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( L )</th>
<th>( H )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_2 )</td>
<td>( q_1 q_2 + \rho \sigma )</td>
<td>( q_1(1 - q_2) - \rho \sigma )</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>( q_1(1 - q_2) + \rho \sigma )</td>
<td>( q_2 )</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>( (1 - q_1)q_2 - \rho \sigma )</td>
<td>( (1 - q_1)(1 - q_2) + \rho \sigma )</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>( 1 - q_1 )</td>
<td>( 1 - q_2 )</td>
</tr>
</tbody>
</table>

- distribution is parametrized by \((q_1, q_2, \rho)\),
- the std dev is \( \sigma = \sqrt{q_1(1 - q_1)}\sqrt{q_2(1 - q_2)} \).
q₁ and q₂ are publicly known.

the agent privately knows θ₁.

the principal privately knows ρ.

to “close the model”: ρ ∼ F on [ρ, ¯ρ].

all this information is common knowledge.
but what is privacy?

- many axiomatically distinct/competing features.

- what’s the micro foundation for its demand:
  - intrinsic: moral/ethical and hence evolutionary approach,
  - instrumental: fine as long as it not used against me monetarily.

- how to think about its manifestation:
  - is it a type, and hence exogenous,
  - is it a social norm, and hence endogenous.

- might help to
  - clarify the privacy paradox,
  - inform the parameters of new anti-trust laws.
notable omissions from the discussion

- theoretically
  - multidimensional attributes,
  - equilib model of demand & supply of privacy in data economy.
- data, privacy, and authoritative repression
  - identification is typically at the core of a fascist project,
  - big debate in india around biometrics for national id.
- change in the structure of human interactions
  - online dating,
  - role of local communities,
  - incentivizing good behavior through a ”social credit system”,
  - political correctness and wokeness: everything is online.